



UNIVERSITÀ
DEGLI STUDI
DI TRIESTE



Dipartimento di
Ingegneria
e Architettura

Corso di misure meccaniche, termiche e collaudi

Prof. Rodolfo Taccani

Prof. Lucia Parussini

Prof. Marco Bogar

a.a.2023-2024

Outline

- Regressione dei dati

Regressione dei dati

Molti problemi ingegneristici e scientifici riguardano la determinazione di una relazione tra un insieme di variabili.

In molte situazioni, c'è una singola variabile di risposta Y , chiamata anche **variabile dipendente**, che dipende dal valore di un insieme di variabili di input, chiamate anche **variabili indipendenti**, x_1, \dots, x_r .

Il tipo più semplice di relazione tra la variabile dipendente Y e le variabili di input x_1, \dots, x_r è una relazione lineare. Si potrebbe scrivere, per alcune costanti $\beta_0, \beta_1, \dots, \beta_r$ l'equazione

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r$$

Se questa fosse la relazione tra Y e gli $x_i, i = 1, \dots, r$, allora sarebbe possibile (una volta appresi i β_i) prevedere esattamente la risposta per qualsiasi insieme di valori di input.

Regressione dei dati

Tuttavia, in pratica, l'equazione precedente è valida introducendo un errore casuale.
La relazione esplicita è

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r + e$$

dove e , errore casuale, si assume essere una variabile casuale con media 0, o equivalentemente

$$E[Y|\mathbf{x}] = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r$$

dove $\mathbf{x} = (x_1, \dots, x_r)$ è l'insieme delle variabili indipendenti e $E[Y|\mathbf{x}]$ è la risposta attesa dati gli input \mathbf{x} .

L'equazione è chiamata **equazione di regressione lineare**.

Le quantità $\beta_0, \beta_1, \dots, \beta_r$ sono chiamate **coefficienti di regressione**.

Regressione dei dati

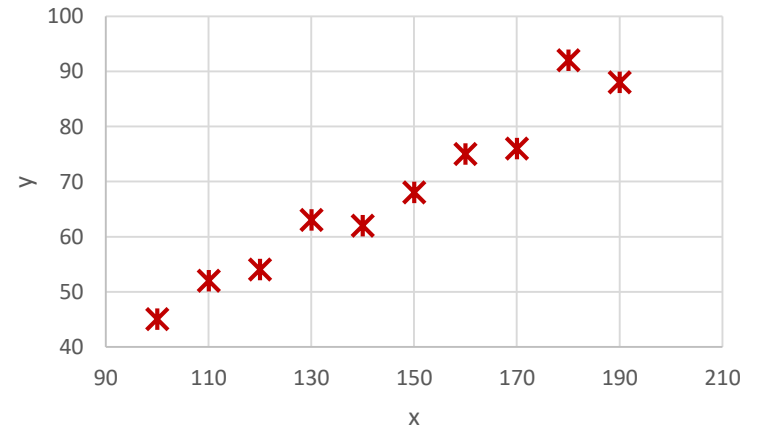
Regressione semplice

Un'equazione di regressione che contiene una sola variabile indipendente - cioè $r = 1$ - è chiamata equazione di regressione semplice:

$$Y = \alpha + \beta x + e$$

ESEMPIO: Consideriamo le seguenti 10 coppie di dati (x_i, y_i) , $i = 1, \dots, 10$, che mettono in relazione y , il risultato percentuale di un esperimento di laboratorio, a x , la temperatura alla quale l'esperimento è stato eseguito.

i	x_i	y_i	i	x_i	y_i
1	100	45	6	150	68
2	110	52	7	160	75
3	120	54	8	170	76
4	130	63	9	180	92
5	140	62	10	190	88



Regressione dei dati

Stimatori ai minimi quadrati dei parametri di regressione

Se A è lo stimatore di α e B di β , allora lo stimatore della risposta corrispondente alla variabile di ingresso x_i sarebbe $A + Bx_i$.

Se la risposta effettiva è Y_i , le quantità $(Y_i - A - Bx_i)_{i=1,\dots,n}$, rappresentano le differenze tra le risposte effettive e le risposte stimate.

$(Y_i - A - Bx_i)_{i=1,\dots,n}$ sono chiamati **residui**.

La somma dei quadrati dei residui è data da

$$SS_R = \sum_{i=1}^n (Y_i - A - Bx_i)^2$$

A e B sono gli stimatori di α e β che **minimizzano** SS_R : **stimatori ai minimi quadrati**.

Regressione dei dati

Stimatori ai minimi quadrati dei parametri di regressione

Per determinare gli stimatori di α e β , che minimizzano SS_R , differenziamo SS_R prima rispetto ad A e B :

$$\frac{\partial SS_R}{\partial A} = -2 \sum_{i=1}^n (Y_i - A - Bx_i)$$

$$\frac{\partial SS_R}{\partial B} = -2 \sum_{i=1}^n x_i (Y_i - A - Bx_i)$$

ponendo le derivate pari a 0:

$$\sum_{i=1}^n (Y_i - A - Bx_i) = 0$$

$$\sum_{i=1}^n x_i (Y_i - A - Bx_i) = 0$$

Quindi otteniamo le equazioni:

$$\sum_{i=1}^n Y_i = nA + B \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i Y_i = A \sum_{i=1}^n x_i + B \sum_{i=1}^n x_i^2$$

note come equazioni normali.

Regressione dei dati

Stimatori ai minimi quadrati dei parametri di regressione

$$\frac{1}{n} \sum_{i=1}^n Y_i = A + B \frac{1}{n} \sum_{i=1}^n x_i \quad \rightarrow \quad \bar{Y} = A + B\bar{x} \quad \rightarrow \quad A = \bar{Y} - B\bar{x}$$

$$\sum_{i=1}^n x_i Y_i = A \sum_{i=1}^n x_i + B \sum_{i=1}^n x_i^2 \quad \rightarrow \quad \sum_{i=1}^n x_i Y_i = (\bar{Y} - B\bar{x})n\bar{x} + B \sum_{i=1}^n x_i^2$$

$$\rightarrow \quad B \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) = \sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y} \quad \rightarrow \quad B = \frac{\sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

Regressione dei dati

Stimatori ai minimi quadrati dei parametri di regressione

Quindi gli stimatori ai minimi quadrati di α e β corrispondenti all'insieme dei dati $(x_i, Y_i)_{i=1, \dots, n}$ sono, rispettivamente,

$$A = \bar{Y} - B\bar{x}$$

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

La linea retta $A + Bx$ è chiamata **retta dei minimi quadrati** o **retta di regressione**.

Regressione dei dati

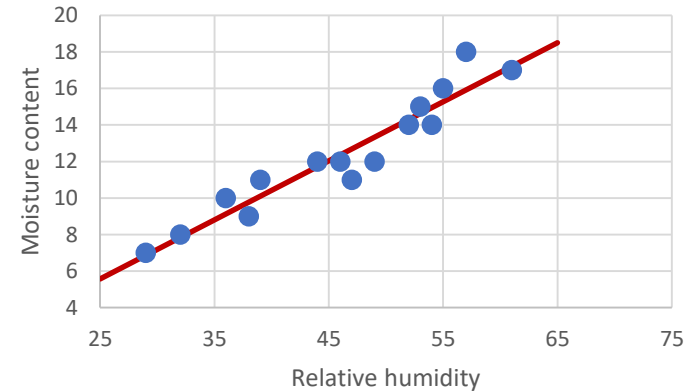
Stimatori ai minimi quadrati dei parametri di regressione

ESEMPIO: La materia prima utilizzata nella produzione di una certa fibra sintetica è conservata in un luogo senza controllo dell'umidità. Le misurazioni dell'umidità relativa nel luogo di stoccaggio e il contenuto di umidità di un campione della materia prima sono stati presi per 15 giorni con i seguenti dati (in percentuale) risultanti.

Relative humidity	46	53	29	61	36	39	47	49	52	38	55	32	57	54	44
Moisture content	12	15	7	17	10	11	11	12	14	9	16	8	18	14	12

$$A = \bar{Y} - B\bar{x} = 12.4 - 46.1B = -2.510$$

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^{15} (x_i - 46.1)Y_i}{\sum_{i=1}^{15} x_i^2 - 15 \cdot 46.1^2} = \frac{\sum_{i=1}^{15} (x_i - 46.1)Y_i}{\sum_{i=1}^{15} x_i^2 - 15 \cdot 46.1^2} = 0.323$$



Regressione dei dati

Distribuzione degli stimatori

Solitamente si assume che gli errori casuali siano variabili casuali normali indipendenti, con media 0 e varianza σ^2 :

$$Y = \alpha + \beta x + e$$

con

$$e \sim N(0, \sigma^2)$$

Quindi, supponiamo che se Y_i è la risposta corrispondente al valore di ingresso x_i , allora Y_1, \dots, Y_n , sono variabili casuali indipendenti e

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2)$$

Regressione dei dati

Distribuzione degli stimatori

Quando le Y_i sono variabili casuali normali, gli stimatori ai minimi quadrati sono anche gli **stimatori di massima verosimiglianza**. Per verificare questa osservazione, si noti che la densità congiunta di Y_1, \dots, Y_n è data da

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = \prod_{i=1}^n f_{Y_i}(y_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \frac{(Y_i - \alpha - \beta x_i)^2}{\sigma^2}} = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\sum_{i=1}^n \frac{1}{2} \frac{(Y_i - \alpha - \beta x_i)^2}{\sigma^2}}$$

Di conseguenza, gli stimatori di massima verosimiglianza di α e β sono precisamente i valori di α e β che minimizzano

$$\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2$$

cioè, sono gli stimatori ai minimi quadrati.

Regressione dei dati

Distribuzione degli stimatori

Se Y_1, \dots, Y_n , sono variabili casuali indipendenti con

$$Y_i \sim N(\alpha + \beta x_i, \sigma^2)$$

A e B sono entrambi una combinazione lineare delle variabili casuali normali indipendenti $Y_i, i = 1, \dots, n$, quindi sono anche normalmente distribuiti.

Regressione dei dati

Distribuzione degli stimatori

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

La media e la varianza di B sono calcolate come segue:

$$E[B] = \frac{\sum_{i=1}^n (x_i - \bar{x}) E[Y_i]}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) (\alpha + \beta x_i)}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{\alpha \sum_{i=1}^n (x_i - \bar{x}) + \beta \sum_{i=1}^n x_i (x_i - \bar{x})}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

Essendo $\sum_{i=1}^n (x_i - \bar{x}) = 0$

$$E[B] = \beta \frac{\sum_{i=1}^n x_i (x_i - \bar{x})}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \beta \frac{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \beta$$

Quindi B è uno stimatore non distorto (unbiased estimator) di β .

Regressione dei dati

Distribuzione degli stimatori

$$\begin{aligned} \text{Var}[B] &= \frac{\text{Var}[\sum_{i=1}^n (x_i - \bar{x}) Y_i]}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}[Y_i]}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)^2} = \sigma^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)^2} \\ &= \sigma^2 \frac{\sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)^2} = \sigma^2 \frac{\sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n \bar{x}^2}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)^2} = \sigma^2 \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2)^2} = \\ &= \sigma^2 \frac{1}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \end{aligned}$$

Regressione dei dati

Distribuzione degli stimatori

$$A = \bar{Y} - B\bar{x} = \frac{1}{n} \sum_{i=1}^n Y_i - B\bar{x}$$

Calcoliamo media e varianza dello stimatore:

$$E[A] = \frac{1}{n} \sum_{i=1}^n E[Y_i] - E[B]\bar{x} = \frac{1}{n} \sum_{i=1}^n (\alpha + \beta x_i) - \beta\bar{x} = \frac{n\alpha}{n} + \beta \frac{1}{n} \sum_{i=1}^n x_i - \beta\bar{x} = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha$$

Quindi A è uno stimatore non distorto (unbiased estimator) di α .

Regressione dei dati

Distribuzione degli stimatori

$$\begin{aligned} \text{Var}[A] &= \text{Var}\left[\frac{1}{n}\sum_{i=1}^n Y_i - B\bar{x}\right] = \frac{1}{n^2}\sum_{i=1}^n \text{Var}[Y_i] + \bar{x}^2 \text{Var}[B] = \frac{\sigma^2}{n} + \sigma^2 \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \\ &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right) = \sigma^2 \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2 - n\bar{x}^2}{n(\sum_{i=1}^n x_i^2 - n\bar{x}^2)} = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n(\sum_{i=1}^n x_i^2 - n\bar{x}^2)} \end{aligned}$$

Regressione dei dati

Distribuzione degli stimatori

Quindi A e B sono stimatori normalmente distribuiti con media e la varianza:

$$E[A] = \alpha$$
$$Var[A] = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n(\sum_{i=1}^n x_i^2 - n\bar{x}^2)}$$

$$E[B] = \beta$$
$$Var[B] = \sigma^2 \frac{1}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

Regressione dei dati

Distribuzione degli stimatori

Se indichiamo con

$$S_{xY} = \sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y}) = \sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y}$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

$$S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$$

allora:

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} = \frac{S_{xY}}{S_{xx}}$$

$$A = \bar{Y} - B\bar{x}$$

$$SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}}$$

Regressione dei dati

Distribuzione degli stimatori

Gli stimatori ai minimi quadrati di α e β

$$A = \bar{Y} - B\bar{x} \qquad B = \frac{S_{xY}}{S_{xx}}$$

sono distribuiti come segue

$$A \sim N\left(\alpha, \frac{\sigma^2 \sum_{i=1}^n x_i^2}{nS_{xx}}\right) \qquad B \sim N(\beta, \sigma^2/S_{xx})$$

La somma dei quadrati degli errori $SS_e = \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2$ è

$$\frac{SS_e}{\sigma^2} \sim \chi_{n-2}^2$$

e SS_e è indipendente da A e B .

Regressione dei dati

Distribuzione degli stimatori

Se

$$\frac{SS_e}{\sigma^2} \sim \chi_{n-2}^2$$

quindi

$$E \left[\frac{SS_e}{\sigma^2} \right] = n - 2 \rightarrow E \left[\frac{SS_e}{n - 2} \right] = \sigma^2 \rightarrow \frac{SS_R}{n - 2} = \sigma^2$$

Quindi $SS_R/(n - 2)$ è uno stimatore non distorto di σ^2 .

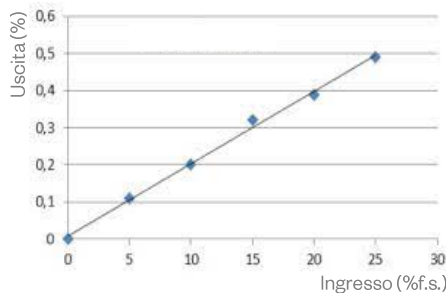
Regressione dei dati

Taratura statica

q_i valore campione

q_o valore misurato

q_i	q_o
$q_{i,1}$	$q_{o,1}$
$q_{i,2}$	$q_{o,2}$
\vdots	\vdots
$q_{i,N}$	$q_{o,N}$



Equazione della retta interpolante $q_o = mq_i + b$

Interpolazione ai minimi quadrati: m e b minimizzano la somma dei quadrati delle distanze $\sum_{j=1}^N (q_{o,j} - mq_{i,j} - b)^2$

$$m = \frac{N \sum q_i q_o - \sum q_i \sum q_o}{N \sum q_i^2 - (\sum q_i)^2}$$

$$b = \frac{\sum q_o \sum q_i^2 - \sum q_i q_o \sum q_i}{N \sum q_i^2 - (\sum q_i)^2}$$

$$s_m^2 = \frac{N s_{q_o}^2}{N \sum q_i^2 - (\sum q_i)^2} \quad s_b^2 = \frac{s_{q_o}^2 \sum q_i^2}{N \sum q_i^2 - (\sum q_i)^2}$$

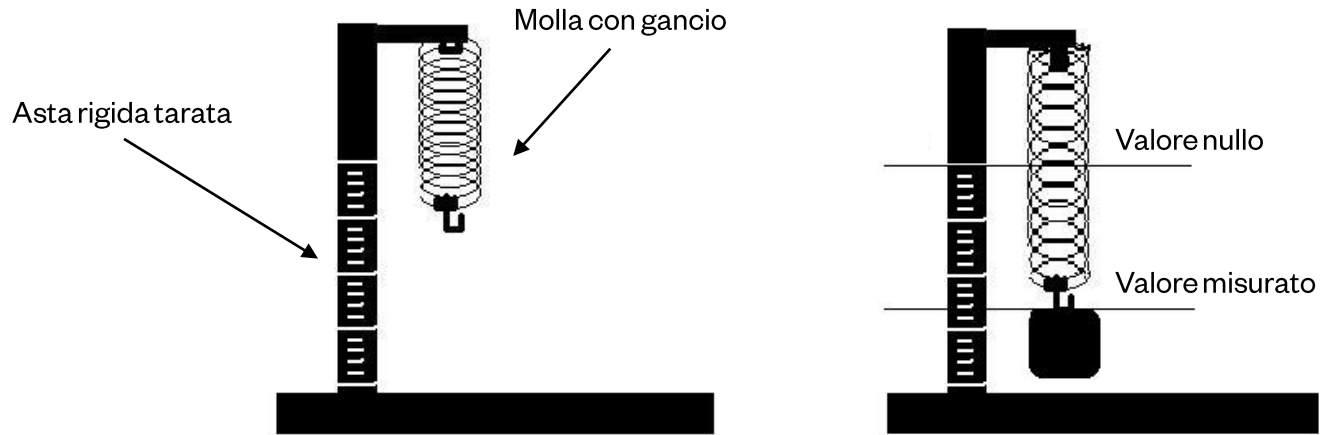
$$s_{q_o}^2 = \frac{1}{N-2} \sum (mq_i + b - q_o)^2$$

$$s_{q_i} = \frac{s_{q_o}}{m}$$

Regressione dei dati

Taratura statica

Esempio di taratura di un dinamometro a molla



Regressione dei dati

q_i	q_o
0	0.1
1	1.25
2	1.98
3	3.01
4	3.95
5	5.06
6	6.09
7	6.9
8	8.15
9	9.02
10	10.1

Equazione della retta di regressione $q_o = 1.004q_i - 0.0737$

Con un livello di confidenza del 95% ovvero un'ampiezza dell'intervallo di $\pm 2\sigma$

$$m = 1.004 \pm 0.022$$

$$b = -0.0737 \pm 0.0226$$

con: $s_m^2 = 0.000121$ e $s_b^2 = 0.00013456$

$$s_{q_o}^2 = \frac{1}{9} \sum (1.004q_i - 0.0737 - q_o)^2 = 0.01054$$

Da cui

$$s_{q_i} = 0.102283$$

Quando il dinamometro rileva $q_o = 1.98$, la misura è $q_i = 2.0455 \pm 0.3068$ se consideriamo un livello di confidenza del 99.7%, con un fattore di copertura pari a 3.

Regressione dei dati

Coefficiente di determinazione e coefficiente di correlazione campionaria

Supponiamo di voler misurare la quantità di variazione nell'insieme dei valori di risposta Y_1, \dots, Y_n corrispondente all'insieme dei valori di ingresso x_1, \dots, x_n . Una misura standard in statistica della quantità di variazione in un insieme di valori Y_1, \dots, Y_n è data dalla quantità:

$$S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

La variazione nei valori di Y_i deriva da due fattori:

- i valori di input x_i sono diversi
- ciascuna delle variabili di risposta Y_i ha varianza σ^2

Regressione dei dati

Coefficiente di determinazione e coefficiente di correlazione campionaria

$SS_R = \sum_{i=1}^n (Y_i - A - Bx_i)^2$ misura la quantità rimanente di variazione nei valori di risposta dopo che i diversi valori di input sono stati presi in considerazione.

$S_{YY} - SS_R$ rappresenta la quantità di variazione nelle variabili di risposta che è spiegata dai diversi valori di input.

Il coefficiente di determinazione, definito come

$$R^2 = \frac{S_{YY} - SS_R}{S_{YY}} = 1 - \frac{SS_R}{S_{YY}}$$

rappresenta la proporzione della variazione nelle variabili di risposta che è spiegata dai diversi valori di input.

Regressione dei dati

Coefficiente di determinazione e coefficiente di correlazione campionaria

R^2 è spesso usato come indicatore di quanto bene il modello di regressione si adatti ai dati.

$$0 \leq R^2 \leq 1$$

Un valore di R^2 vicino a 1 indica che la maggior parte della variazione dei dati di risposta è spiegata dai diversi valori di input

- buon adattamento del modello ai dati

Un valore di R^2 vicino a 0 indica che poco della variazione è spiegata dai diversi valori di input

- cattivo adattamento del modello ai dati

Regressione dei dati

Coefficiente di determinazione e coefficiente di correlazione campionaria

Il coefficiente di correlazione campionaria r dell'insieme di coppie di dati $(x_i, Y_i), i = 1, \dots, n$, è definito da

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{S_{xY}}{\sqrt{S_{xx}S_{YY}}}$$

Essendo $SS_R = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}}$ vediamo che

$$r^2 = \frac{S_{xY}^2}{S_{xx}S_{YY}} = \frac{S_{xx}S_{YY} - S_{xx}SS_R}{S_{xx}S_{YY}} = 1 - \frac{SS_R}{S_{YY}} = R^2 \quad \rightarrow \quad |r| = \sqrt{R^2}$$

e quindi, a parte il suo segno che indica se è positivo o negativo, il coefficiente di correlazione campionaria è uguale alla radice quadrata del coefficiente di determinazione.

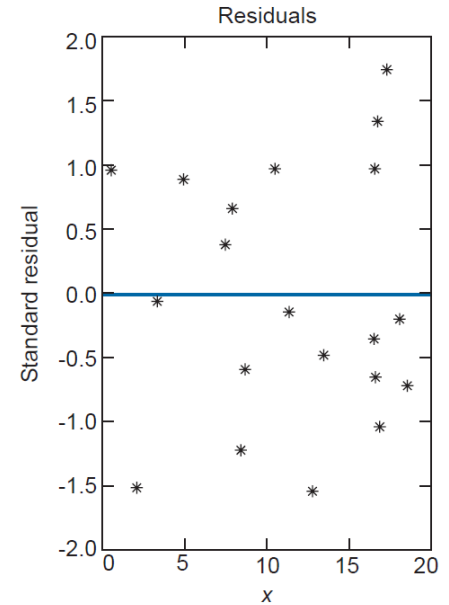
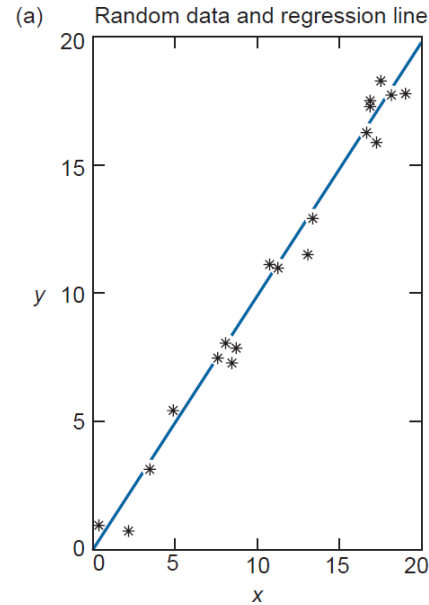
Regressione dei dati

Analisi dei residui: valutazione del modello

residui standardizzati

$$\frac{Y_i - (A + Bx_i)}{\sqrt{SS_R/(n - 2)}}$$

Il modello, sia dal diagramma di dispersione che dalla natura casuale dei suoi residui standardizzati, sembra adattarsi abbastanza bene al modello a linea retta.



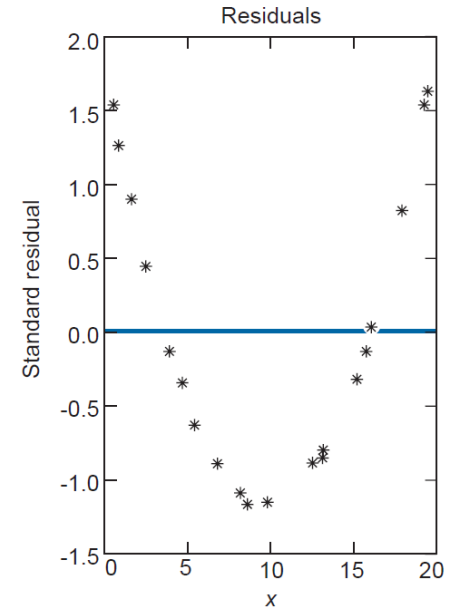
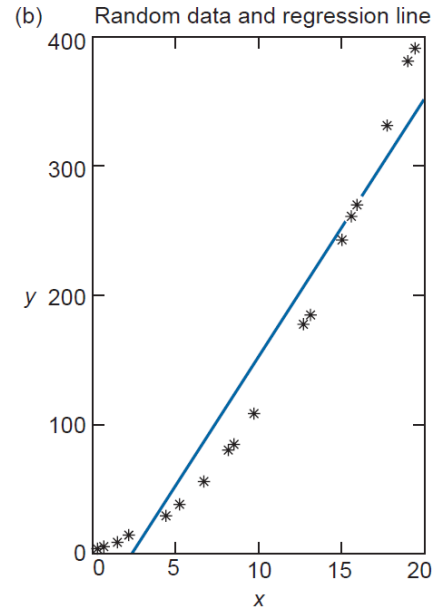
Regressione dei dati

Analisi dei residui: valutazione del modello

residui standardizzati

$$\frac{Y_i - (A + Bx_i)}{\sqrt{SS_R/(n - 2)}}$$

Il diagramma dei residui mostra che i residui prima diminuiscono e poi aumentano all'aumentare del livello di ingresso. Questo spesso significa che sono necessari termini di ordine superiore (non solo lineari) per descrivere la relazione tra l'input e la risposta. Si vede anche dal diagramma di dispersione in questo caso.



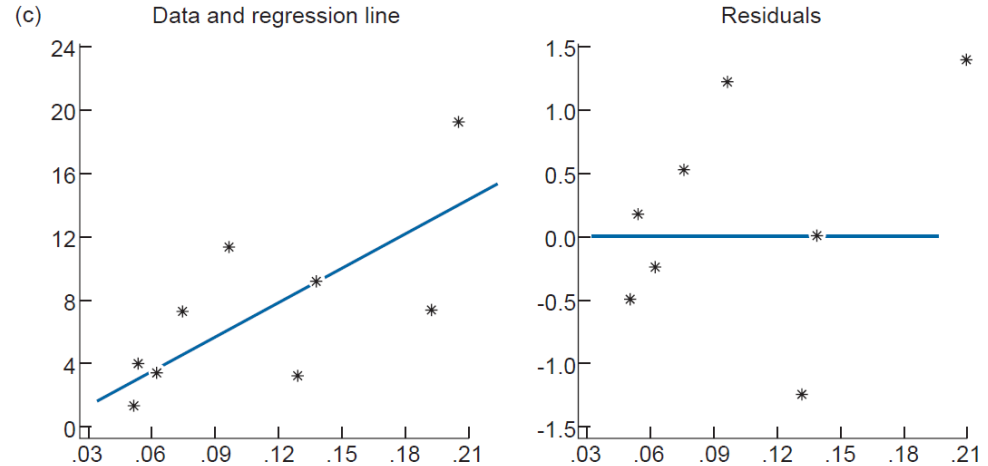
Regressione dei dati

Analisi dei residui: valutazione del modello

residui standardizzati

$$\frac{Y_i - (A + Bx_i)}{\sqrt{SS_R/(n - 2)}}$$

Il diagramma dei residui standardizzati mostra che il valore assoluto dei residui, e quindi i loro quadrati, aumentano all'aumentare del livello di input. Questo spesso indica che la varianza della risposta non è costante ma, piuttosto, aumenta con il livello di input.

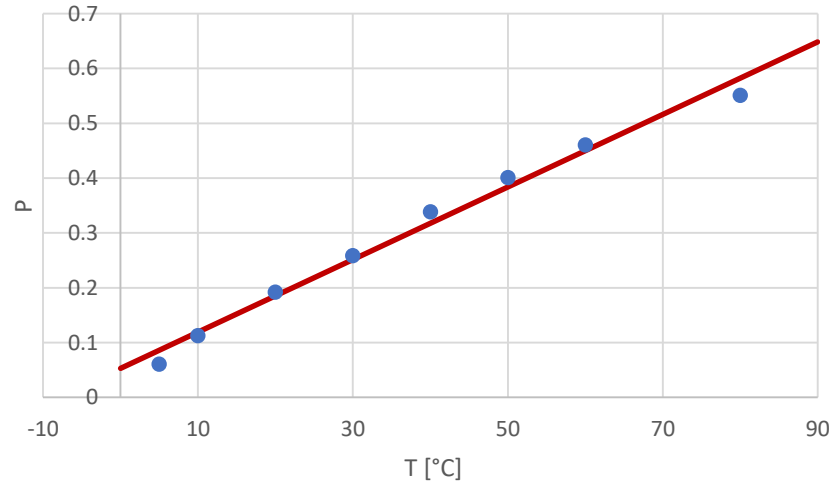


Regressione dei dati

Linearizzazione

ESEMPIO: La seguente tabella fornisce le percentuali di una sostanza chimica che sono state consumate quando un esperimento è stato eseguito a varie temperature (in gradi celsius). Usala per stimare la percentuale della sostanza chimica che verrebbe consumata se l'esperimento venisse eseguito a 350 gradi.

Temperature	Percentage
5°	0.061
10°	0.113
20°	0.192
30°	0.259
40°	0.339
50°	0.401
60°	0.461
80°	0.551



Regressione dei dati

Linearizzazione

ESEMPIO:

Sia $P(x)$ la percentuale della sostanza chimica che si consuma quando l'esperimento viene eseguito a $10x$ gradi. Anche se un grafico di $P(x)$ sembra approssimativamente lineare, possiamo migliorare l'adattamento considerando una relazione non lineare tra x e $P(x)$.

In particolare, consideriamo una relazione della forma

$$1 - P(x) \approx c(1 - d)^x$$

Cioè, supponiamo che la percentuale della sostanza chimica che sopravvive ad un esperimento condotto alla temperatura x diminuisce approssimativamente ad un tasso esponenziale quando x aumenta. Prendendo i logaritmi, il precedente può essere scritto come

$$\ln(1 - P(x)) \approx \ln(c) + x \ln(1 - d)$$

Così, impostando

$$Y = -\ln(1 - P(x))$$

$$\alpha = -\ln c$$

$$\beta = -\ln(1 - d)$$

otteniamo

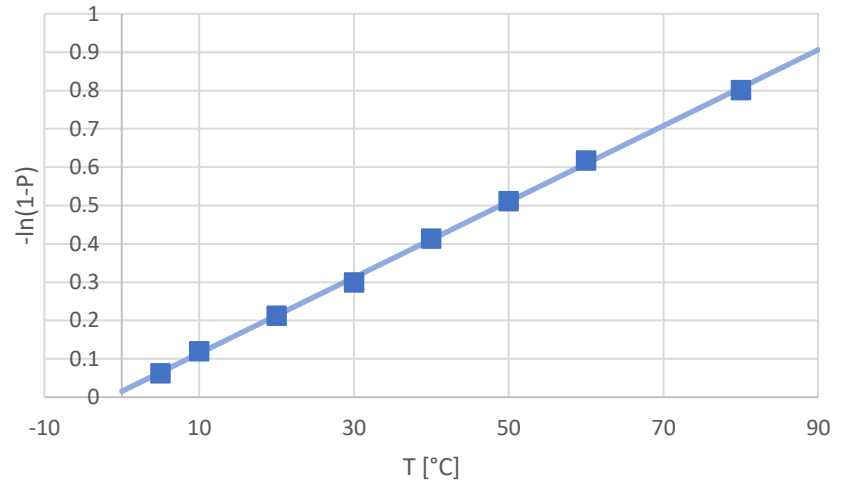
$$Y = \alpha + \beta x + e$$

Regressione dei dati

Linearizzazione

ESEMPIO:

Temperature	$-\ln(1 - \text{Percentage})$
5°	0.063
10°	0.120
20°	0.213
30°	0.300
40°	0.414
50°	0.512
60°	0.618
80°	0.801



$$A = 0.0154$$

$$B = 0.0099$$

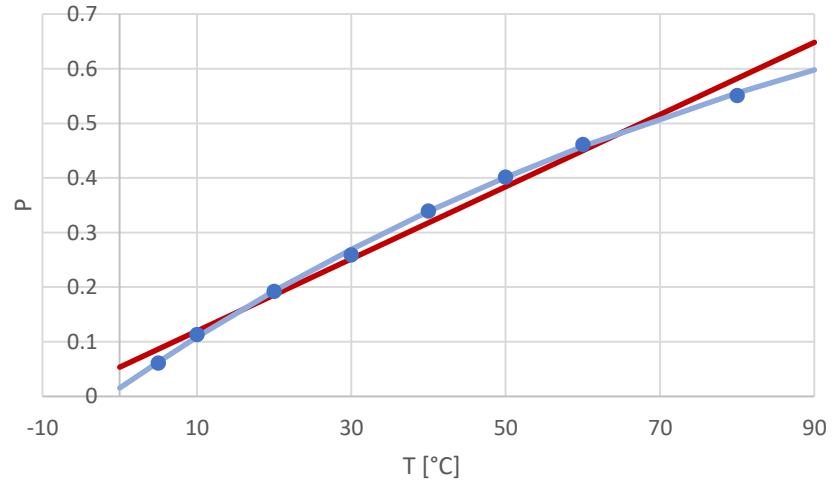
Regressione dei dati

Linearizzazione

ESEMPIO: Nella variabile originale

$$\hat{c} = e^{-A} = 0.9847$$
$$1 - \hat{d} = e^{-B} = 0.9901$$
$$\hat{P} = 1 - 0.9847(0.9901)^x$$

T	P	\hat{P}	$P - \hat{P}$
5	0.061	0.063	-0.002
10	0.113	0.109	0.004
20	0.192	0.193	-0.001
30	0.259	0.269	-0.010
40	0.339	0.339	0.000
50	0.401	0.401	0.000
60	0.461	0.458	0.003
80	0.551	0.556	-0.005



Regressione dei dati

Regressione polinomiale

In situazioni in cui la relazione funzionale tra la risposta Y e la variabile indipendente variabile x non può essere adeguatamente approssimata da una relazione lineare, potremmo provare ad adattare alla serie di dati una relazione funzionale della forma

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_r x^r + e$$

dove $\beta_0, \beta_1, \dots, \beta_r$ sono coefficienti di regressione che devono essere stimati.

Se l'insieme dei dati è costituito dalle n coppie $(x_i, Y_i), i = 1, \dots, n$, allora gli stimatori ai minimi quadrati di $\beta_0, \beta_1, \dots, \beta_r$ - chiamiamoli B_0, B_1, \dots, B_r - sono quei valori che minimizzano

$$SS_R = \sum_{i=1}^n (Y_i - B_0 - B_1 x_i - B_2 x_i^2 - \dots - B_r x_i^r)^2$$

Regressione dei dati

Regressione polinomiale

Ponendo le derivate parziali di SS_R rispetto a B_0, B_1, \dots, B_r , uguale a 0, gli stimatori ai minimi quadrati B_0, B_1, \dots, B_r soddisfano il seguente insieme di $r + 1$ equazioni lineari chiamate **equazioni normali**:

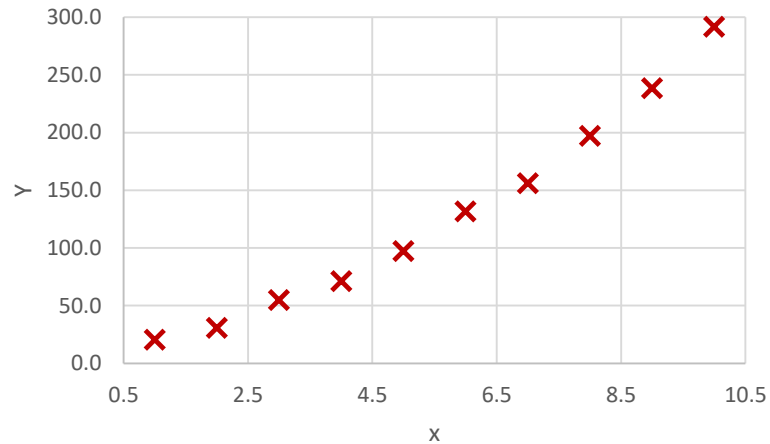
$$\begin{aligned} \sum_{i=1}^n Y_i &= B_0 n + B_1 \sum_{i=1}^n x_i + B_2 \sum_{i=1}^n x_i^2 + \dots + B_r \sum_{i=1}^n x_i^r \\ \sum_{i=1}^n x_i Y_i &= B_0 \sum_{i=1}^n x_i + B_1 \sum_{i=1}^n x_i^2 + B_2 \sum_{i=1}^n x_i^3 + \dots + B_r \sum_{i=1}^n x_i^{r+1} \\ \sum_{i=1}^n x_i^2 Y_i &= B_0 \sum_{i=1}^n x_i^2 + B_1 \sum_{i=1}^n x_i^3 + B_2 \sum_{i=1}^n x_i^4 + \dots + B_r \sum_{i=1}^n x_i^{r+2} \\ &\vdots = \quad \quad \quad \vdots \quad + \quad \quad \quad \vdots \quad + \quad \quad \quad \vdots \quad + \dots + \quad \quad \quad \vdots \\ \sum_{i=1}^n x_i^r Y_i &= B_0 \sum_{i=1}^n x_i^r + B_1 \sum_{i=1}^n x_i^{r+1} + B_2 \sum_{i=1}^n x_i^{r+2} + \dots + B_r \sum_{i=1}^n x_i^{2r} \end{aligned}$$

Regressione dei dati

Regressione polinomiale

ESEMPIO: Adatta un polinomio ai seguenti dati:

x	Y
1	20.6
2	30.8
3	55.0
4	71.4
5	97.3
6	131.8
7	156.3
8	197.3
9	238.7
10	291.7



Regressione dei dati

Regressione polinomiale

ESEMPIO:

Un grafico di questi dati indica che una relazione quadratica

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + e$$

è sufficiente. Siccome:

$$\sum_{i=1}^n x_i = 55 \quad \sum_{i=1}^n x_i^2 = 385 \quad \sum_{i=1}^n x_i^3 = 3025 \quad \sum_{i=1}^n x_i^4 = 25333$$

$$\sum_{i=1}^n Y_i = 1291.1 \quad \sum_{i=1}^n x_i Y_i = 9549.3 \quad \sum_{i=1}^n x_i^2 Y_i = 77758.9$$

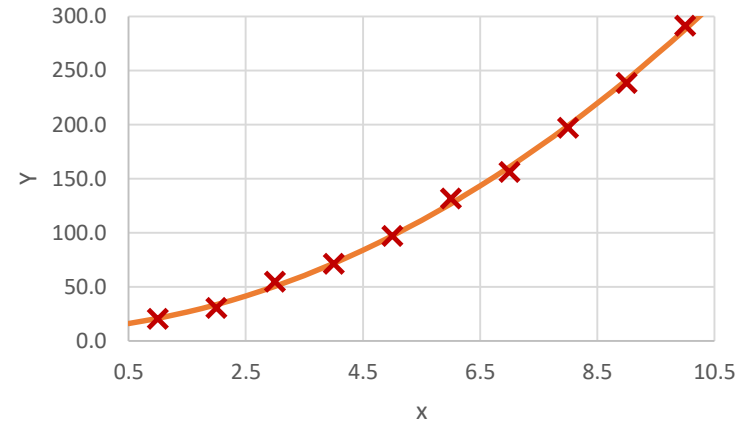
gli stimatori ai minimi quadrati sono le soluzioni del seguente sistema di equazioni:

$$\begin{aligned} 1291.1 &= 10B_0 + 55B_1 + 385B_2 \\ 9549.3 &= 55B_0 + 385B_1 + 3025B_2 \\ 77758.9 &= 385B_0 + 3025B_1 + 25333B_2 \end{aligned}$$

Da cui

$$B_0 = 12.59326 \quad B_1 = 6.326172 \quad B_2 = 2.122818$$

$$Y = 12.59 + 6.33x + 2.12x^2$$



Regressione dei dati

Regressione lineare multipla

Nella maggior parte delle applicazioni, una situazione tipica è quella in cui c'è un insieme di k variabili di input e la risposta Y è legata ad esse dalla relazione

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + e$$

dove $x_j, j = 1, \dots, k$, è il livello della j -esima variabile di ingresso ed e è un errore casuale che assumiamo sia normalmente distribuito con media 0 e varianza (costante) σ^2 . Cioè, le Y_i sono collegate ai livelli di ingresso attraverso la:

$$E[Y_i] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

Se B_0, B_1, \dots, B_k denotano gli stimatori di $\beta_0, \beta_1, \dots, \beta_k$, allora

$$SS_R = \sum_{i=1}^n (Y_i - B_0 - B_1 x_{i1} - B_2 x_{i2} - \dots - B_k x_{ik})^2$$

Regressione dei dati

Regressione lineare multipla

Per determinare gli stimatori dei minimi quadrati, poniamo le derivate parziali di SS_R rispetto a B_0, B_1, \dots, B_k uguale a 0. Otteniamo $k + 1$ equazioni

$$\begin{aligned}\sum_{i=1}^n (Y_i - B_0 - B_1 x_{i1} - B_2 x_{i2} - \dots - B_k x_{ik}) &= 0 \\ \sum_{i=1}^n x_{i1} (Y_i - B_0 - B_1 x_{i1} - B_2 x_{i2} - \dots - B_k x_{ik}) &= 0 \\ \sum_{i=1}^n x_{i2} (Y_i - B_0 - B_1 x_{i1} - B_2 x_{i2} - \dots - B_k x_{ik}) &= 0 \\ &\vdots \\ \sum_{i=1}^n x_{ik} (Y_i - B_0 - B_1 x_{i1} - B_2 x_{i2} - \dots - B_k x_{ik}) &= 0\end{aligned}$$

Regressione dei dati

Regressione lineare multipla

Le equazioni normali sono:

$$\begin{aligned}\sum_{i=1}^n Y_i &= nB_0 + B_1 \sum_{i=1}^n x_{i1} + B_2 \sum_{i=1}^n x_{i2} + \cdots + B_k \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} Y_i &= B_0 \sum_{i=1}^n x_{i1} + B_1 \sum_{i=1}^n x_{i1}^2 + B_2 \sum_{i=1}^n x_{i1} x_{i2} + \cdots + B_k \sum_{i=1}^n x_{i1} x_{ik} \\ &\vdots \\ \sum_{i=1}^n x_{ik} Y_i &= B_0 \sum_{i=1}^n x_{ik} + B_1 \sum_{i=1}^n x_{ik} x_{i1} + B_2 \sum_{i=1}^n x_{ik} x_{i2} + \cdots + B_k \sum_{i=1}^n x_{ik}^2\end{aligned}$$

Regressione dei dati

Regressione lineare multipla

In notazione matriciale:

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

con \mathbf{Y} matrice $n \times 1$, \mathbf{X} matrice $n \times p$, $\boldsymbol{\beta}$ matrice $p \times 1$, ed \mathbf{e} matrice $n \times 1$ dove $p \equiv k + 1$.

Il modello di regressione multipla è

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Regressione dei dati

Regressione lineare multipla

Sia

$$\mathbf{B} = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix}$$

la matrice degli stimatori ai minimi quadrati, allora le equazioni normali possono essere scritte come:

$$\mathbf{X}^T \mathbf{X} \mathbf{B} = \mathbf{X}^T \mathbf{Y}$$

Da cui

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Regressione dei dati

Regressione di processi gaussiani (GPR)

I modelli GPR sono una classe di modelli machine learning non parametrici comunemente utilizzati per modellare dati spaziali e serie temporali.

I modelli GPR hanno diversi vantaggi:

- funzionano bene su piccoli insiemi di dati
- forniscono misure di incertezza sulle previsioni.

I modelli GPR consentono una regressione non lineare dei dati.

Regressione dei dati

Regressione di processi gaussiani (GPR)

Il modello GPR non è parametrico (cioè non è limitato da una forma funzionale).

Invece che calcolare la distribuzione di probabilità dei parametri di una funzione specifica, il modello GPR calcola la distribuzione di probabilità su tutte le funzioni ammissibili che si adattano ai dati.

Specifichiamo una prior (sullo spazio delle funzioni), calcoliamo la posterior usando i dati di allenamento, e calcoliamo la distribuzione predittiva a posteriori sui punti di interesse.

Ci sono diverse librerie che implementano efficientemente il modello (ad esempio scikit-learn, Gpytorch, GPy).

Regressione dei dati

Regressione di processi gaussiani (GPR)

Definizioni

Una **distribuzione Gaussiana multivariata** è definita da un vettore medio $\boldsymbol{\mu}$ e una matrice di covarianza $\boldsymbol{\Sigma}$. La media $\boldsymbol{\mu}$ si riferisce alla media di ciascuna variabile casuale X_i e la matrice di covarianza $\boldsymbol{\Sigma}$ fornisce la covarianza tra ciascuna delle variabili casuali X_i e X_j :

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1j}^2 & \cdots & \sigma_{1n}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2j}^2 & \cdots & \sigma_{2n}^2 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \sigma_{i1}^2 & \sigma_{i2}^2 & \cdots & \sigma_{ij}^2 & \cdots & \sigma_{in}^2 \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \sigma_{n2}^2 & \cdots & \sigma_{nj}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix}$$

Intuitivamente, la matrice di covarianza generalizza la varianza a più dimensioni e la diagonale è costituita dalla varianza di ciascuna variabile casuale X_i .

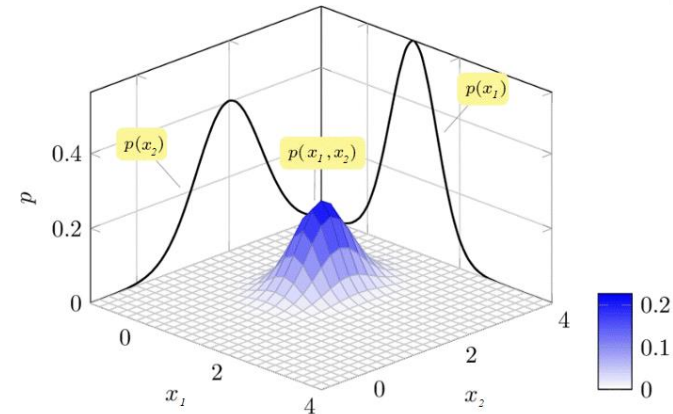
Regressione dei dati

Regressione di processi gaussiani (GPR)

Definizioni

Sia $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ una coppia di variabili aleatorie con distribuzione di probabilità congiunta $p(x_1, x_2)$ gaussiana con media $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ e covarianza $\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.

$$p(x_1, x_2) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$



Esempio di una distribuzione gaussiana biviata.

Regressione dei dati

Regressione di processi gaussiani (GPR)

Definizioni

Si possono derivare le seguenti utili formule in forma chiusa per le probabilità marginali e condizionali:

$$p(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_2 = N(\mu_1, \Sigma_{11})$$

$$p(x_2) = \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 = N(\mu_2, \Sigma_{22})$$

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} = N(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Regressione dei dati

Regressione di processi gaussiani (GPR)

Definizioni

I processi aleatori sono un'estensione del concetto di variabile aleatoria.

Si definisce processo stocastico una famiglia di variabili aleatorie $X(\mathbf{s}, \omega)$, $\mathbf{s} \in \mathcal{D} \subset \mathbb{R}^d$, definite su uno spazio campione Ω con $\omega \in \Omega$, e che assumono valori in un insieme definito *spazio degli stati del processo*. Un processo stocastico è quindi un insieme di funzioni che evolvono nello spazio \mathcal{D} (le cosiddette *funzioni campione* o *realizzazioni*), ognuna delle quali è associata ad un determinato elemento dello spazio campione, così che il risultato di un esperimento casuale corrisponde di fatto all'estrazione di una di queste funzioni.

Il processo stocastico è una funzione stocastica specificata dalle sue distribuzioni congiunte dimensionalmente finite:

$$F(y_1, y_2, \dots, y_n; \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n) = P(X(\mathbf{s}_1) \leq y_1, \dots, X(\mathbf{s}_n) \leq y_n)$$

per ogni n finito e ogni insieme di punti $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ in \mathcal{D} .

Regressione dei dati

Regressione di processi gaussiani (GPR)

Definizioni

In teoria delle probabilità un **processo gaussiano** è un processo stocastico $X(\mathbf{s})$ tale che prendendo un qualsiasi numero finito di variabili aleatorie, dalla collezione che forma il processo aleatorio stesso, esse hanno una distribuzione di probabilità congiunta gaussiana.

Un processo gaussiano è specificato interamente dalla sua media $\mu(\mathbf{s})$ e dalla covarianza $k(\mathbf{s}, \mathbf{s}') = \text{Cov}(X(\mathbf{s}); X(\mathbf{s}'))$ e viene indicato nel modo seguente:

$$X(\mathbf{s}) \sim N(\mu(\mathbf{s}), k(\mathbf{s}, \mathbf{s}'))$$

Ha la proprietà che per ogni insieme finito di punti $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$

$$\mathbf{x} \equiv (X(\mathbf{s}_1), X(\mathbf{s}_2), \dots, X(\mathbf{s}_n))^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

con $\Sigma_{ij} = k(\mathbf{s}_i, \mathbf{s}_j)$.

Regressione dei dati

Regressione di processi gaussiani (GPR)

Generazione dei campioni dato $\mu(\cdot)$ e $k(\cdot, \cdot)$ (esempio 1d):

- 1) Consideriamo l'insieme di n punti $\mathbf{s} = (s_1, s_2, \dots, s_n)^T$
- 2) Calcoliamo la matrice di covarianza Σ nei punti
- 3) Calcoliamo la decomposizione di Cholesky di $\Sigma = \mathbf{L}\mathbf{L}^T$
- 4) Generiamo il campione

$$x(\mathbf{s}) = \mu(\mathbf{s}) + \mathbf{L}^T \cdot \text{randn}(n)$$

dove $\text{randn}(n)$ restituisce n scalari tratti da una distribuzione normale standard

Regressione dei dati

Regressione di processi gaussiani (GPR)

Definizioni

Per l'esistenza di un **processo gaussiano** con media e covarianza prescritte è sufficiente assicurare che k sia definita positiva. In questo caso la distribuzione ha densità:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

Una funzione k è definita positiva se per ogni insieme finito di punti $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ in \mathcal{D} la matrice di covarianza

$$\boldsymbol{\Sigma} = \begin{bmatrix} k(\mathbf{s}_1, \mathbf{s}_1) & k(\mathbf{s}_1, \mathbf{s}_2) & \cdots & k(\mathbf{s}_1, \mathbf{s}_n) \\ k(\mathbf{s}_2, \mathbf{s}_1) & k(\mathbf{s}_2, \mathbf{s}_2) & \cdots & k(\mathbf{s}_2, \mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{s}_n, \mathbf{s}_1) & k(\mathbf{s}_n, \mathbf{s}_2) & \cdots & k(\mathbf{s}_n, \mathbf{s}_n) \end{bmatrix}$$

è semi-definita positiva: $\mathbf{z}^T \boldsymbol{\Sigma} \mathbf{z} \geq 0$ per ogni vettore a valori reali \mathbf{z} .

Regressione dei dati

Regressione di processi gaussiani (GPR)

Nel modello GPR, la risposta $z(\mathbf{s})$ è considerata una realizzazione del processo Gaussiano multivariato $Z(\mathbf{s})$:

$$Z(\mathbf{s}) = m(\mathbf{s}) + Y(\mathbf{s})$$

con $m(\mathbf{s}) = \mathbf{h}(\mathbf{s})\boldsymbol{\beta}$ è una funzione di regressione deterministica, costruita dai dati osservati, e $Y(\mathbf{s})$ è un processo Gaussiano, costruito sui residui, con media nulla e funzione di covarianza $k(\mathbf{s}_i, \mathbf{s}_j) = \sigma^2 r(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$.

σ^2 è un fattore di scala, chiamato VARIANZA del PROCESSO

$r(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta})$ è una funzione positiva con parametri $\boldsymbol{\theta}$, chiamata funzione di CORRELAZIONE

$\boldsymbol{\theta}$ sono detti IPER- PARAMETRI

Regressione dei dati

Regressione di processi gaussiani (GPR)

Se $m(\mathbf{s}) = 0$, il modello GPR è detto **SIMPLE KRIGING**.

Se $m(\mathbf{s}) = \beta$ con β costante, il modello GPR è detto **ORDINARY KRIGING**.

Se $m(\mathbf{s}) = \mathbf{h}(\mathbf{s})\boldsymbol{\beta}$ con $\mathbf{h}(\mathbf{s}) = (h_1(\mathbf{s}), \dots, h_p(\mathbf{s}))^T$ p funzioni scelte e $\boldsymbol{\beta}$ vettore di p coefficienti incogniti, il modello GPR è detto **UNIVERSAL KRIGING**.

Regressione dei dati

Regressione di processi gaussiani (GPR)

Indichiamo con $\mathbf{z}^{(n)}$ i valori osservati di $z(\mathbf{s})$ in n punti noti $\hat{\mathcal{D}} = (\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n)^T \subset \mathcal{D}$.

In molti casi, non abbiamo accesso diretto alla funzione da approssimare ma solo a una sua versione rumorosa. Consideriamo questo caso più generale, assumendo un rumore di osservazione gaussiano indipendente con media zero e varianza $\sigma_\epsilon^2(\mathbf{s})$. Questo è solitamente indicato come nugget effect. Quindi, $\mathbf{z}^{(n)}$ sono realizzazioni del vettore gaussiano $\mathbf{Z}^{(n)} = Z(\hat{\mathcal{D}}) + \mathbf{E}^{(n)}$, dove $Z(\hat{\mathcal{D}})$ è il processo gaussiano $Z(\mathbf{s})$ nei punti $\hat{\mathcal{D}}$ e $\mathbf{E}^{(n)} = (\sigma_\epsilon(\mathbf{s}_1)E_1, \dots, \sigma_\epsilon(\mathbf{s}_n)E_n)^T$ è il rumore bianco con $E_{i=1, \dots, n}$ indipendenti e identicamente distribuite rispetto a una distribuzione gaussiana con media zero e varianza uno.

Per semplicità assumiamo un modello ordinary kriging, dove $m(\mathbf{s}) = h(\mathbf{s})\beta$ con $h(\mathbf{s}) = 1$ e $p = 1$, e σ_ϵ costante.

Regressione dei dati

Regressione di processi gaussiani (GPR)

Usiamo le informazioni contenute in $\mathbf{Z}^{(n)}$ per prevedere $Z(\mathbf{s})$ considerando la distribuzione congiunta di $Z(\mathbf{s})$ e $\mathbf{Z}^{(n)}$:

$$\begin{pmatrix} Z(\mathbf{s}) \\ \mathbf{Z}^{(n)} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{h}(\mathbf{s})\boldsymbol{\beta} \\ \mathbf{H}\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} k(\mathbf{s}, \mathbf{s}) & \mathbf{k}^T(\mathbf{s}) \\ \mathbf{k}(\mathbf{s}) & \boldsymbol{\Sigma} + \sigma_\epsilon^2 \mathbf{I} \end{pmatrix} \right)$$

dove

$\mathbf{H} = \mathbf{h}(\hat{\mathcal{D}})$ matrice del modello $n \times p$

$\boldsymbol{\Sigma}$ matrice di covarianza $n \times n$ tra i punti osservati $\hat{\mathcal{D}}$

$\mathbf{k}(\mathbf{s})$ vettore di covarianza di dimensione n tra i punti da predire \mathbf{s} e i punti osservati $\hat{\mathcal{D}}$

Regressione dei dati

Regressione di processi gaussiani (GPR)

La distribuzione condizionale

$$p(Z(\mathbf{s})|\mathbf{Z}^{(n)}) = \frac{p(Z(\mathbf{s}), \mathbf{Z}^{(n)})}{p(\mathbf{Z}^{(n)})}$$

è gaussiana con media e varianza:

$$\hat{m}_Z(\mathbf{s}) = \mathbf{h}(\mathbf{s})\boldsymbol{\beta} + \mathbf{k}^T(\mathbf{s})(\boldsymbol{\Sigma} + \sigma_\epsilon^2\mathbf{I})^{-1}(\mathbf{z}^{(n)} - \mathbf{H}\boldsymbol{\beta})$$

$$\hat{s}_Z^2(\mathbf{s}) = k(\mathbf{s}, \mathbf{s}) - \mathbf{k}^T(\mathbf{s})(\boldsymbol{\Sigma} + \sigma_\epsilon^2\mathbf{I})^{-1}\mathbf{k}(\mathbf{s})$$

Regressione dei dati

Regressione di processi gaussiani (GPR)

$$[Z|\mathbf{Z}^{(n)}; \boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}] \sim N(\hat{m}_Z, \hat{s}_Z^2)$$

$$\text{prior } Z = GP(m(\cdot), k(\cdot, \cdot))$$

$$\text{likelihood } p(\mathbf{Z}^{(n)}|Z; \boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}) = N(Z, \sigma_\epsilon^2 \mathbf{I})$$

$$\text{marginal likelihood } p(\mathbf{Z}^{(n)}; \boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}) = \int p(\mathbf{Z}^{(n)}|Z; \boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}) p(Z; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) dZ$$

$$p(Z|\mathbf{Z}^{(n)}; \boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}) = \frac{p(\mathbf{Z}^{(n)}|Z; \boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}) p(Z; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})}{p(\mathbf{Z}^{(n)}; \boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta})}$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}$$

Regressione dei dati

Regressione di processi gaussiani (GPR)

Per stimare $\boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}$ un metodo molto popolare è la Stima della Massima Verosimiglianza (Maximum Likelihood Estimation MLE).

L'assunzione di distribuzione normale multivariata per $\mathbf{Z}^{(n)}$ porta alla seguente likelihood:

$$p(\mathbf{Z}^{(n)} | \mathbf{Z}; \boldsymbol{\beta}, \sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma} + \sigma_\epsilon^2 \mathbf{I}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{z}^{(n)} - \mathbf{H}\boldsymbol{\beta})^T (\boldsymbol{\Sigma} + \sigma_\epsilon^2 \mathbf{I})^{-1} (\mathbf{z}^{(n)} - \mathbf{H}\boldsymbol{\beta})\right)$$

Dato

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T (\boldsymbol{\Sigma} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{H})^{-1} \mathbf{H}^T (\boldsymbol{\Sigma} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{z}^{(n)}$$

che è la MLE di $\boldsymbol{\beta}$ corrispondente alla sua stima generalizzata ai minimi quadrati, la stima di $\sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}$ si ottiene minimizzando

$$\mathcal{L}_{LME}(\sigma^2, \sigma_\epsilon^2, \boldsymbol{\theta}) = (\mathbf{z}^{(n)} - \mathbf{H}\boldsymbol{\beta})^T (\boldsymbol{\Sigma} + \sigma_\epsilon^2 \mathbf{I})^{-1} (\mathbf{z}^{(n)} - \mathbf{H}\boldsymbol{\beta}) + \log(|\boldsymbol{\Sigma} + \sigma_\epsilon^2 \mathbf{I}|)$$

che è l'opposto della log-likelihood a meno di una costante.

Regressione dei dati

Regressione di processi gaussiani (GPR)

Come scegliere la *prior* $Z = GP(m(\cdot), k(\cdot, \cdot))$:

La scelta del kernel o covarianza è molto importante e specifica come *credi* che *sia* la funzione latente Z .

Il kernel maggiormente utilizzato è il Gaussian Kernel o RBF o exponential quadratic, definito dalla funzione:

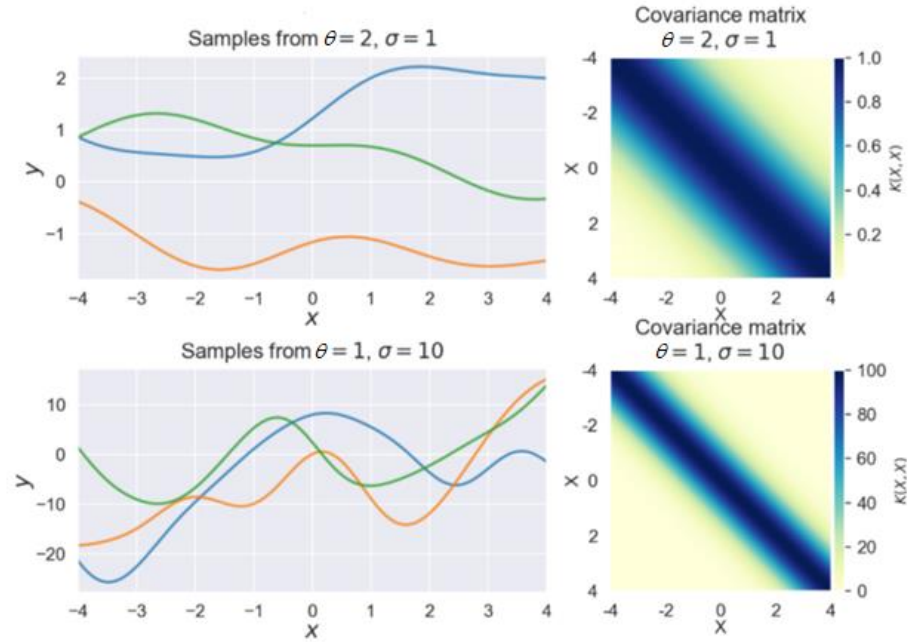
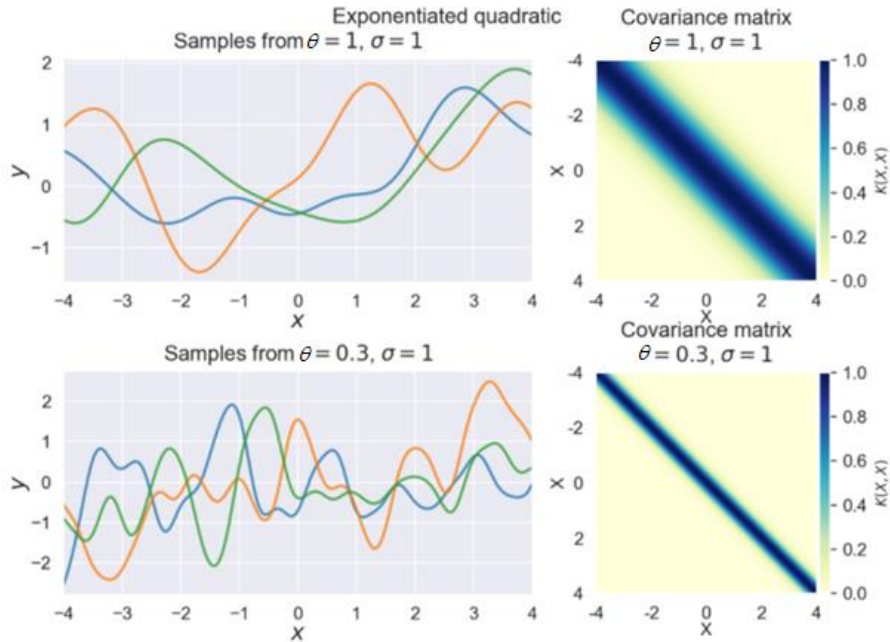
$$k(x, x') = \sigma^2 \exp\left(-\frac{1}{2\theta^2} \|x - x'\|^2\right)$$

θ è la scala di lunghezza

σ^2 è la varianza del GP

Regressione dei dati

Regressione di processi gaussiani (GPR)



Regressione dei dati

Regressione di processi gaussiani (GPR)

Exponential quadratic kernel

$$k(x, x') = \sigma^2 \exp\left(-\frac{1}{2\theta^2} \|x - x'\|^2\right)$$

θ è la scala di lunghezza: descrive quanto velocemente la correlazione tra due osservazioni x e x' si riduce quanto più sono lontane.

Un alto θ risulta in una funzione regolare.

Un basso θ risulta in una funzione irregolare.

Il parametro σ^2 , varianza del GP, controlla l'ampiezza verticale della funzione.

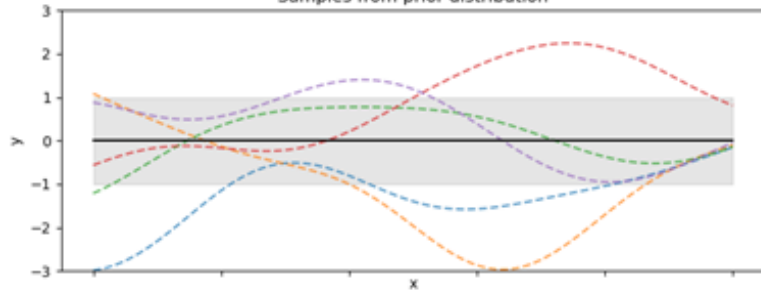
Sono stati sviluppati un gran numero di kernel sia stazionari che instazionari: Matern 3/2, Matern 5/2, exponential, cubic, periodic, Gibbs, neural network...

Regressione dei dati

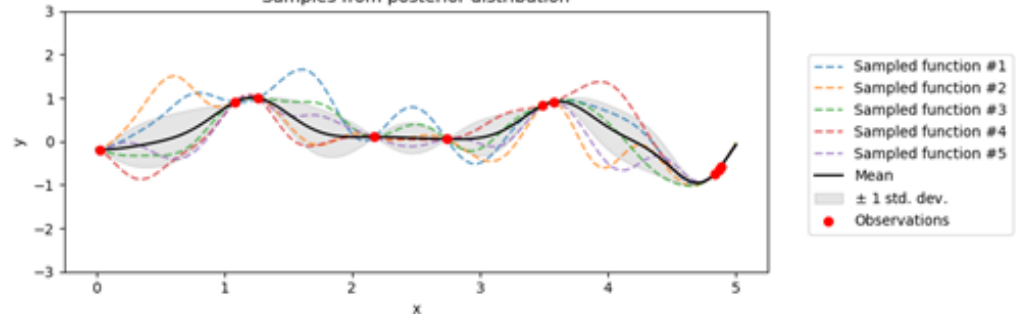
Regressione di processi gaussiani (GPR)

Radial Basis Function kernel

Samples from prior distribution

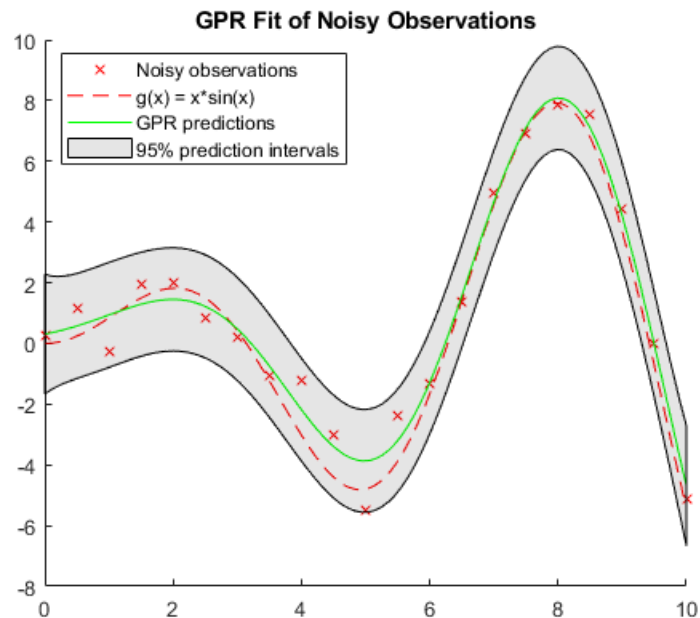
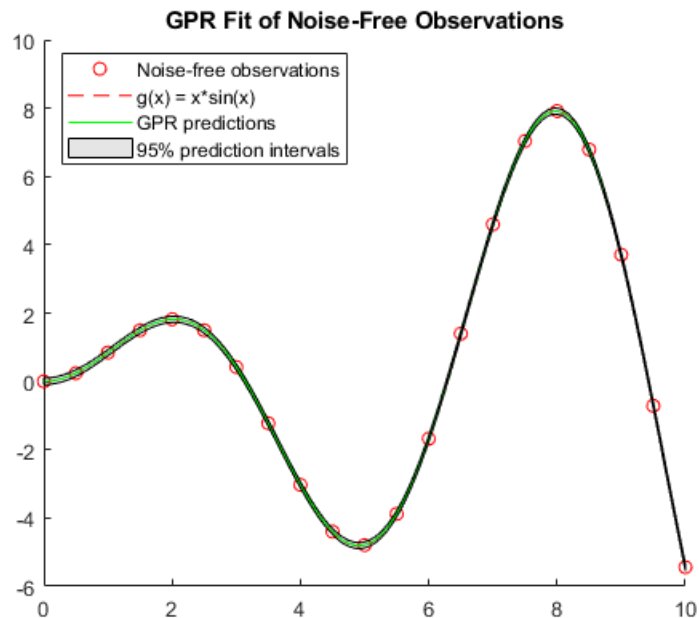


Samples from posterior distribution



Regressione dei dati

Regressione di processi gaussiani (GPR)





UNIVERSITÀ
DEGLI STUDI
DI TRIESTE



Dipartimento di
**Ingegneria
e Architettura**